

IMC-41U

TITLE

5 **SYNONYM EXTENSION OF SEARCH QUERIES WITH
VALIDATION**

RELATED APPLICATIONS

10 This is a continuation-in-pat of copending U.S. Patent Application
Serial No. 60/199,659 filed April 25, 2000 and copending U.S. Patent
Application Serial No. 60/199,920 filed April 26, 2000. This application is
related to copending U.S. Patent Application Serial No. 60/199,658 filed April
25, 2000 and copending U.S. Patent Application Serial No. 60/199,921 filed
15 April 26, 2000. This application is also related to copending U.S. Patent
Application Serial No. 09/541,192 filed April 3, 2000, which is a continuation
application of copending US Patent Application Serial No. 09/345,547, filed
June 30,1999 which is a continuation-in-part of copending U.S. Patent
Application Serial No. 09/321,804 filed May 27, 1999. These applications are
20 herewith incorporated herein by reference.

FIELD OF THE INVENTION

25 This invention relates to computer based search systems, and
particularly to narrowing searches for the user's convenience.

BACKGROUND OF THE INVENTION

Usually computer-based document-search processors use keywords. The result of a keyword search is often an enormous amount of information, the majority of which is irrelevant to user's requirements.

5 The precision of a search can be increased, if the user formulates the query as a problem, for example, "produce aluminum layer", or "heat water", or "oxidize silicon wafer". But such formulations severely decrease the quantity of found information because the search will not find expressions such as "form Al layer" that are similar to "produce aluminum layer", or
10 expressions such as "increase temperature of water" that are similar to "heat water", or expressions such as "perform oxidation of silicon substrate" that are similar to "oxidize silicon wafer".

 It is possible to provide a complete and more exact search with help of
15 a dictionary of search synonyms e.g. in the form of Al layer = produce aluminum layer..., etc. But estimates show that volume of entries in such a dictionary would be huge – more than 10^7 expressions to describe different problems in just technical fields. Expansion into other (nontechnical) disciplines will lead to multiple expansion of this dictionary (up to 10^{11}
20 expressions).

An object of the invention is to improve search systems.

SUMMARY OF EMBODIMENTS OF THE INVENTION

25

An embodiment of the invention involves expanding the user query with help of two synonym dictionaries-actions and object, and then validating every result from the obtained queries with help of a Subject-Action-Object Knowledge Database (SAO KB), containing fields with subjects, actions,

objects, and "main parts of objects" extracted from the object. The SAO KB is prepared from natural language texts with the help of a semantic processor such as that disclosed in US Patent No. 6,167,370.

5 These and other embodiments, objects, and advantages of the invention will become evident from the following description of exemplary embodiments when read in light of the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

10

Figure 1 is a block diagram of a system using a software program and embodying the invention.

15 Figure 2 is a flow chart illustrating the operation of the program embodying the invention.

Figure 3 is a more detailed flowchart illustrating the operation of the program.

20 Figure 4 is a view of a screen in a monitor depicting the program and inviting entry of a query.

Figure 5 is a view of a screen in a monitor depicting the program and inviting entry of a query.

25

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The following are incorporated herein by reference:

I. System and on-line information service presently available at www.cobrain.com and the publicly available user manual therefor.

5 II. The software product presently marketed by Invention Machine Corporation of Boston, Massachusetts, USA, under it's trademark "KNOWLEDGIST" and the publicly available user manual therefor.

10 III. US Patent No. 6,167,370.

IV. U.S. Patent Application Serial No. 09/541,182 filed April 3, 2000.

15 V. The software product presently marketed by Invention Machine Corporation of Boston, Massachusetts, USA under its Trademark "TECHOPTIMIZER" and the publicly available user manual therefor.

VI. U.S. Patent No. 5,901,068.

20 In Figure 1 a preferred software system and method embodying the invention is in the form of a program. The program resides in a personal computer 12 that includes a CPU 14, a monitor 16, a keyboard/mouse 18, and a printer 20 and is in the form of a program. The program may be stored on a portable disk and inserted in a disk reader slot 22 or on a fixed disc in the computer or on a ROM. According to another embodiment the program
25 resides on a server and the user accesses the program with a standard communication port 23 over a communications network via LAN, WAN, or the Internet. The port 23 also serves for accessing information from databases on the Internet. Computer 12 can be conventional and be of any suitable

make or brand. A printer 20 provides a hard copy of a session where desired. Other peripherals and modem/network interfaces are provided as desired.

Figure 2 is a flowchart illustrating the operation of an embodiment.

5 Here a user's query 50 passes to a query analysis and extension phase 60 to produce a list of queries 70. A queries validation unit 80 using an SAO KB validates the query list to produce a desired list of validated queries 90 for search in external documents.

10 A more detailed diagram of query extension appears in Figure 3. Here, the program invites a user to enter a query 110 on a semantic search screen of a monitor 16 as shown in Figure 4. The user's query data 120 passes to query analysis module 130. Query analysis module 130 divides the user's query data 120 into action data 150 and object data 160. The simplest method of such
15 division relies on list of verbs. For example, we can extract action "heat" from an initial query: "heat aluminum layer", because "heat" is present in the list of verbs; the remainder of query: "aluminum layer" is then recognized as an object.

20 Action data 150 are transmitted to module 180 of action expansion (or action expansion module). Action expansion is accomplished on the basis of an action dictionary 140, containing the six parts:

- List of verbs divided into groups, containing the verbs with similar sense (heat-warm, produce-create-generate, etc.);
- 25 • List of "verb-nouns" expressions synonymous with other verb (heat-increase temperature-rise temperature, etc.)
- List of "verbsA" including the verbs -perform, carry out, realize, and other verbs with similar sense;

- List of “nouns” including the following groups – “verb - relevant verbal noun” (heat-heating; produce-production, etc.)
- List of “verbsB” including the verbs –produce, create, form, and other verbs of similar sense;
- 5 • List of “participle2” including the following groups – “verb - relevant participle2” (heat-heated; produce-produced, etc.).

The action expansion module 180 forms four lists from action module 150 and actions dictionary 140:

- 10 • List of verbs relevant to action in module 200;
- List of “verbsA - verbal noun” expressions relevant to action in module 210;
- List of “verbsB - participle2” expressions relevant to action in module 220;
- 15 • List of “verb - noun” expressions relevant to action in module 230.

For example, module 180 takes the action “heat” and forms the following lists:

- List of verbs relevant to action in module 200 – “heat, warm”;
- 20 List of “verbsA - verbal noun” expressions relevant to action in module 210 – “perform – heating”, “carry out – heating”, “realize – heating”, etc.;
- List of “verbsB - participle2” expressions relevant to action in module 220 – “produce – heated”, “create – heated”, “form – heated”, and etc.
- 25 List of “verb - noun” expressions relevant to action in module 230 – “increase - temperature”, rise – temperature”, and etc.

The object data 160 pass to the module 190 of object expansion (or object expansion module 190). Object expansion is accomplished on basis of a

dictionary of object synonyms 170, containing groups of objects having similar sense (aluminum layer - Al layer, laser radiation – coherent radiation, etc.).

5 The object expansion module 190 forms lists of expressions synonymous with the objects in module 240 from the objects in module 160 and a dictionary of object synonyms in module 170. For example, object expansion in module 190 uses the object “aluminum layer” to form the following list: aluminum layer, Al layer, aluminium layer, etc.

10

All the lists in modules 200, 210, 220, 230, and 240 are transmitted into validation module 250. A validation module 250 forms search queries from the lists in modules 200, 210, 220, 230, and 240 and accomplishes the search in an SAO KB 260 according to these queries.

15

The SAO KB 260 is prepared from natural language texts with help of a semantic processor as described in the aforementioned US Patent Application Serial No. 09/345,547 filed April 3, 2000 (Reference IV above) as well as in US Patent No. 6,167,370 (Reference III above). The SAO KB 260 contains the following fields: subjects, actions, objects, and “main parts of objects” extracted from the object. For example, the semantic processor converts the sentence: “A thin aluminum layer is heated by reflected laser radiation” into following fields of the SAO KB 260:

20

25

Subject – “reflected laser radiation”;
Action – “heat”;
Object – “thin aluminum layer”;
Main part of object – “aluminum layer”;

The volume of entries in the SAO KB 260 should be about 10^6 SAOs or more. The natural language texts can belong to various disciplines such as (science, engineering, culture, business, etc.). A customized SAO KB can also be used. Texts for a customized SAO KB should be selected from a
 5 single discipline. If the user query and customized SAO KB belong to similar disciplines, query expansion will be more complete and precise. According to an embodiment the SAO KB 260 is prepared from natural language texts in a specific discipline or a group of related disciplines.

10 Validation module 250 performs search queries according to the following rules. Here the sequence is not relevant although all are performed:

1) [All verbs from list of verbs relevant to action 200 through OR]
 AND [all expressions from list of synonymous to object expressions
 240 through OR];

15 the search of verbs is accomplished in the "action" field of the SAO KB 260, and search of expressions from list of synonymous to object expressions is accomplished in "main part of object" field of SAO KB 260;

2) [All verbs from list of "verbsA" through OR] AND [all nouns
 from list of "verbsA - verbal noun" expressions relevant to action 210
 20 through OR] AND [all expressions from list of synonymous to object
 expressions 240 through OR];

the search of verbs is accomplished in the "action" field of the SAO KB 260, search of verbal nouns from "verbsA - verbal noun" expressions is accomplished in the "main part of object" field of the SAO KB 260; and
 25 search of expressions from list of synonymous to object expressions is accomplished in the "object" field of SAO KB 260;

3) [All verbs from list of “verbsB” through OR] AND [all participles2 from list of “verbsB-participle2” expressions relevant to action 220 through OR] AND [all expressions from list of synonymous to object expressions 240 through OR];

5 the search of verbs is accomplished in the “action” field of the SAO KB 260, the search of participles2 from list of “verbsB-participle2” expressions is accomplished in the “object” field of SAO KB; and the search of expressions from the list of synonymous to object expressions is accomplished in the “main part of object” field of the SAO KB.

10 4) [All verbs from list of “verb - noun” expressions relevant to action 230 through OR] AND [all nouns from list of “verbs - noun” expressions relevant to action 230 through OR] AND [all expressions from list of synonymous to object expressions 240 through OR];

15 the search of verbs is accomplished in the “action” field of the SAO KB 260, the search of nouns from the “verbs - noun” expressions is accomplished in the “main part of object” field of the SAO KB; and search of expressions from list of synonymous to object expressions is accomplished in the “object” field of the SAO KB 260;

20 For example, it is possible to obtain the following search queries from “heat aluminum layer” (for simplification, not all the verbs and expressions from the dictionaries are used):

- 25 • {[heat OR warm] in “action” field of SAO KB} AND {[aluminum layer OR Al layer] in “main part of object” field of SAO KB};

- {[perform OR realize] in “action” field of SAO KB} AND {[heating] in “main part of object” field of SAO KB} AND {[aluminum layer OR Al layer] in “object” field of SAO KB};
- 5 • {[produce OR create] in “action” field of SAO KB} AND {[aluminum layer OR Al layer] in “main part of object” field of SAO KB} AND {[heated] in “object” field of SAO KB};
- {[increase OR rise] in “action” field of SAO KB} AND {[temperature] in “main part of object” field of SAO KB} AND {[aluminum layer OR Al layer] in “object” field of SAO KB}.
- 10

The search queries are used by validation module 250 for searching in SAO KB 260. The module 250 counts the quantity of found SAOs for every query. If an SAO isn't found, the query is considered non-valid. Then the validation module 250 forms the list of validated SAOs 280 comprising the SAOs found according to the above-mentioned queries in SAO KB 260. Module 310 shows the user validated SAOs, as illustrated in Figure 5 which is a view of a screen displaying data from the program.

20

Furthermore, the validation module 250 can form the list of validated AOs (action-object) 270 from the list of validated SAOs 280. For that, module 250 removes subjects from all validated SAOs and all words from objects, except those contained in search queries. A list of validated AOs 270 can be used in module 300 of search (or search module 300) for searching external information sources 290.

25

Two examples (for simplification, using only the actions dictionary; with the volume of an SAO KB= 5×10^6 SAOs) are shown in following table.

User query – form magnetic film. Queries after expansion		Results of search in SAO KB with frequencies			Queries after validation	
Form	Magnetic Film	Form	Magnetic Film	3724	Form	Magnetic Film
Produce	Magnetic Film	Produce	Magnetic Film	262	Produce	Magnetic Film
Obtain	Magnetic Film	Obtain	Magnetic Film	220	Obtain	Magnetic Film
Provide	Magnetic Film	Provide	Magnetic Film	211	Provide	Magnetic Film
Make	Magnetic Film	Make	Magnetic Film	126	Make	Magnetic Film
Grow	Magnetic Film	Grow	Magnetic Film	87	Grow	Magnetic Film
Fabricate	Magnetic Film	Fabricate	Magnetic Film	42	Fabricate	Magnetic Film
Give	Magnetic Film	Give	Magnetic Film	42	Give	Magnetic Film
Create	Magnetic Film	Create	Magnetic Film	24	Create	Magnetic Film
Manufacture	Magnetic Film	Manufacture	Magnetic Film	15	Manufacture	Magnetic Film
Prepare	Magnetic Film	Prepare	Magnetic Film	14	Prepare	Magnetic Film
Generate	Magnetic Film	Generate	Magnetic Film	11	Generate	Magnetic Film
Synthesize	Magnetic Film	Synthesize	Magnetic Film	6	Synthesize	Magnetic Film
Emit	Magnetic Film	Emit	Magnetic Film	0		
Radiate	Magnetic Film	Radiate	Magnetic Film	0		
Give Off	Magnetic Film	Give Off	Magnetic Film	0		
Emanate	Magnetic Film	Emanate	Magnetic Film	0		
Construct	Magnetic Film	Construct	Magnetic Film	0		
Yield	Magnetic Film	Yield	Magnetic Film	0		

Acquire	Magnetic Film	Aequire	Magnetic Film	0		
Derive	Magnetic Film	Derive	Magnetic Film	0		
User query – produce laser radiation. Queries after expansion		Results of search in SAO KB with frequencies			Queries after validation	
Form	Laser Radiation	Emit	Laser Radiation	834	Emit	Laser Radiation
Produce	Laser Radiation	Generate	Laser Radiation	271	Generate	Laser Radiation
Obtain	Laser Radiation	Produce	Laser Radiation	173	Produce	Laser Radiation
Provide	Laser Radiation	Provide	Laser Radiation	81	Provide	Laser Radiation
Make	Laser Radiation	Form	Laser Radiation	19	Form	Laser Radiation
Grow	Laser Radiation	Make	Laser Radiation	10	Make	Laser Radiation
Fabricate	Laser Radiation	Radiate	Laser Radiation	9	Radiate	Laser Radiation
Give	Laser Radiation	Obtain	Laser Radiation	5	Obtain	Laser Radiation
Create	Laser Radiation	Yield	Laser Radiation	5	Yield	Laser Radiation
Manufacture	Laser Radiation	Grow	Laser Radiation	0		
Prepare	Laser Radiation	Fabricate	Laser Radiation	0		
Generate	Laser Radiation	Give	Laser Radiation	0		
Synthesize	Laser Radiation	Create	Laser Radiation	0		
Emit	Laser Radiation	Manufaeture	Laser Radiation	0		
Radiate	Laser Radiation	Prepare	Laser Radiation	0		
Give Off	Laser Radiation	Synthesi ze	Laser Radiation	0		
Emanate	Laser Radiation	Give Off	Laser Radiation	0		
Construct	Laser	Emanate	Laser	0		

	Radiation		Radiation			
Yield	Laser Radiation	Construe t	Laser Radiation	0		
Acquire	Laser Radiation	Aequire	Laser Radiation	0		
Derive	Laser Radiation	Derive	Laser Radiation	0		

Results with zero frequencies are deleted. These examples show that use of validation based on the SAO KB 260 provides an exact and relevant expansion of a user query. At the same time, used dictionaries have acceptable values.

The invention refines a synonym expansion of a user query by comparison of automatically generated set of synonym queries in the form "action - object" with actually existing "action - object" relations of an SAO KB. This results in a set of relevant synonyms for queries to provide exact and complete search results

It will be understood that various other display symbols, emblems, colors, and configurations can be used instead of those disclosed for the exemplary embodiments herein. Also, various improvements and modifications can be made to the herein-disclosed exemplary embodiments without departing from the spirit and scope of the present invention. The system and method according to the inventive principles herein are necessarily not dependent upon the precise exemplary hardware or software architecture disclosed herein.